

Available Data Resources for MT06

Circle the corpora you wish to receive and fax your request to:

Attn: Audrey Le
+1 (301) 670-0939

Arabic Lexicons

LDC2004L02 Buckwalter Arabic Morphological Analyzer Version 2.0

Arabic <-> English Parallel Text

LDC2004E13 UN Arabic English Parallel Text
LDC2004T17 Arabic News Translation Text Part 1
LDC2004T18 Arabic English Parallel News Part 1
LDC2005E46 Arabic Treebank English Translation
LDC2004E72 eTIRR Arabic English News Text

Arabic <-> English Multiple Translation

LDC2003T18 Multiple-Translation Arabic (MTA) Part 1
LDC2005T05 Multiple-Translation Arabic (MTA) Part 2
LDC2006E44 TIDES MT 2004 Arabic evaluation data
LDC2006E39 TIDES MT 2005 Arabic evaluation data

Arabic Treebanks

LDC2005T02 Arabic Treebank: Part 1 v 3.0 (POS with full vocalization + syntactic analysis)
LDC2004T02 Arabic Treebank: Part 2 v 2.0
LDC2005T20 Arabic Treebank: Part 3 (full corpus) v2.0 (MPG + Syntactic Analysis)

Chinese Lexicons

- | | |
|------------|---|
| LDC2002L27 | Chinese-English Translation Lexicon Version 3.0 |
| LDC2005T34 | Chinese <-> English Name Entity Lists (v1.0) |

Chinese <-> English Parallel Text

- | | |
|------------|--|
| LDC2004E12 | UN Chinese English Parallel Text |
| LDC2002E18 | Xinhua Chinese English Parallel News Text Version 1 beta |
| LDC2003E07 | Chinese Treebank English Parallel Corpus |
| LDC2004T08 | Hong Kong Parallel Text |
| LDC2005T10 | Chinese English News Magazine Parallel Text |
| LDC2003E14 | FBIS Multilanguage Texts |
| LDC2005T06 | Chinese News Translation Text Part 1 |

Chinese <-> English Multiple Translation

- | | |
|------------|---|
| LDC2002T01 | Multiple-Translation Chinese Corpus |
| LDC2003T17 | Multiple-Translation Chinese (MTC) Part 2 |
| LDC2004T07 | Multiple-Translation Chinese (MTC) Part 3 |
| LDC2006T04 | Multiple Translation Chinese (MTC) Part 4 |
| LDC2006E43 | TIDES MT 2004 Chinese evaluation data |
| LDC2006E38 | TIDES MT 2005 Chinese evaluation data |

Chinese Treebanks

- | | |
|------------|------------------------------|
| LDC2005T01 | Chinese Treebank 5.0 |
| LDC2005T23 | Chinese Proposition Bank 1.0 |

Monolingual Text

LDC2005T12 English Gigaword Second Edition
LDC2006T02 Arabic Gigaword Second Edition
LDC2005T14 Chinese Gigaword Second Edition

GALE Dry-Run Data

LDC2006E17 GALE dry-run data unsegmented source